

VU Research Portal

Testing geographical framing and substitution effects in spatial choice experiments

Schaafsma, M.; Brouwer, R.

published in

Journal of Choice Modelling
2013

DOI (link to publisher)

[10.1016/j.jocm.2013.04.007](https://doi.org/10.1016/j.jocm.2013.04.007)

document version

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

citation for published version (APA)

Schaafsma, M., & Brouwer, R. (2013). Testing geographical framing and substitution effects in spatial choice experiments. *Journal of Choice Modelling*, 8, 32-48. <https://doi.org/10.1016/j.jocm.2013.04.007>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl



Testing geographical framing and substitution effects in spatial choice experiments

Marije Schaafsma^{a,b,*}, Roy Brouwer^{b,1}

^a Department of Environmental Economics, Institute for Environmental Studies, VU University Amsterdam, Amsterdam, Netherlands

^b Centre for Social and Economic Research on the Global Environment (CSERGE), University of East Anglia, Norwich, UK

ARTICLE INFO

Article history:

Received 8 April 2013

Accepted 22 April 2013

Available online 19 June 2013

Keywords:

Choice experiment

Environmental valuation

Water quality

Choice set size

ABSTRACT

One of the main challenges in modelling spatial choices is the complexity resulting from the availability of multiple alternatives at different geographical scales. This study aims to test geographical framing and substitution effects in stated choice experiments by first increasing and subsequently reducing the geographical scale and associated set of choice alternatives in the experiment. Geographical framing effects are tested by comparing estimated choice models for differently sized choice sets. Testing these framing effects related to choice set size helps to inform decisions on choice set composition. The results indicate that changing the choice set size has little to no effect on preference parameters and estimated WTP values. However, the larger choice set is associated with higher error variance, suggesting higher choice task complexity.

Crown Copyright © 2013 Published by Elsevier Ltd. All rights reserved.

1. Introduction

Under the assumption of perfect information, decision-makers in neo-classical consumer choice theory are expected to evaluate all substitutes simultaneously when making choices. Stated preference (SP) studies for environmental valuation usually focus on single study sites. These studies have been criticised for drawing respondent attention to the single study site only, away from available substitutes and thereby inflating willingness-to-pay (WTP) values. Providing sufficient information about the good under valuation as well as its substitutes is a challenge given, for example, the time limitations of social surveys. Simple textual reminders of substitutes, suggested by the well-known NOAA panel (Arrow et al., 1993) to increase the reliability of SP studies, have been argued to be ineffective (Loomis et al., 1994; Whitehead and Blomquist, 1995; Kotchen and Reiling, 1999). The number of single-site SP studies applying contingent valuation (CV) and making an effort to account for substitution is limited. Brown and Duffield (1995) include the number of alternatives and Pate and Loomis (1997) account for the size (acreage) of possible substitutes. A small number of CV studies include multi-programme scenarios in which different goods are valued simultaneously to test for substitution and complementarity effects (Hoehn and Loomis, 1993; Cummings et al., 1994; Hailu et al., 2000; Payne et al., 2000). Cummings et al. (1994) and Neill (1995) argue that in order to elicit WTP estimates that account for substitution effects, respondents should be asked to value substitutes and study sites simultaneously. Alternatively, surveys should provide at least a good description of available substitutes, using pictures, maps or text.

* Corresponding author at: Centre for Social and Economic Research on the Global Environment (CSERGE), University of East Anglia, Norwich, UK.
Tel.: +44 1603 593747; fax: +44 1603 591327.

E-mail address: m.schaafsma@uea.ac.uk (M. Schaafsma).

¹ Tel.: +31 20 5985608; fax: +31 20 5989553.

Choice experiments (CEs) offer the possibility to present study sites and their substitutes as distinct alternatives in a choice set (Rolfe and Bennett, 2006), thereby encouraging respondents to actively consider all available options. The exclusion of relevant substitutes in the choice set may bias parameter estimates and inflate WTP values (DeShazo et al., 2009). In choice studies across (goods provided at) different locations, the number of available alternative sites may be large. Including all these options in the choice set and asking respondents to evaluate all substitution possibilities may lead to a high task complexity and cognitive burden. In those cases, the regularity assumptions underlying rational decision making (Mas-Colell et al., 1995), especially the assumption of completely informed choices, may be violated. To reduce the complexity of choosing among too many available alternatives, respondents may instead base their choices on heuristics and rules of thumb. The complexity may even invoke different, non-compensatory decision strategies, such as elimination-by-aspects (Tversky, 1972). This paper focuses on the effects of CE design dimensions on choice complexity. Besides the design dimensions, choice complexity will also depend on the involvement and interest of respondents in the survey and its topic.

The dilemma that arises for spatial choice studies is that, on the one hand, the assessment of WTP for environmental goods provided at a certain location may require studying a large area to ensure that all relevant substitutes are taken into account.² On the other hand, as the number of alternatives increases with the spatial scale of the study, so does choice complexity, possibly evoking preference anomalies. Better understanding of choice task complexity and violations of rational choice behaviour in spatial choices will help to design comprehensible surveys without reducing the relevance of the design, and thereby increase the reliability of spatial choice studies. The main objective of this paper is to test if geographical framing of the choice set affects WTP and leads to instability of the observed choices in a spatially-explicit CE study. The framing effects in this study consist of a change in the geographical area covered by the choice set, resulting in differently sized choice sets. In the case study, respondents are asked for their WTP for water quality improvements at eleven lakes in a confined geographical area. The experimental choice set begins with the four alternatives nearest to the respondent's home, which allows the respondents to get acquainted with and learn from the choice tasks. Then, the set is increased to seven alternatives from the total set of eleven lakes. The CE ends with the same four nearest lakes again. The alternatives are hence framed in different spatial contexts. By examining differences in WTP as a result of geographical framing and choice set complexity at the same time, this study provides further empirical evidence of the trade-off in choice set specification between the potential increase in complexity involved in large choice sets, and the importance of controlling for relevant substitutes.

The effect of different choice sets on choice behaviour has been tested before (e.g., DeShazo and Fermo, 2002; Arentz et al., 2003), but never sequentially in a SP spatial multiple site selection context and always in cross-section studies (i.e. between respondents). The case study presented here provides the possibility for a within-respondent test of the effects of the choice set size using panel data. The design of the experiment also allows for the analysis of the net effect of learning and fatigue. Choice and preference stability are tested by estimating a pooled mixed logit model (Brownstone and Train, 1998), performing Swait and Louviere (1993) tests on the attribute and scale parameters, and comparing the WTP values resulting from separately estimated models.

The remainder of the paper is organised as follows. The next section reviews the literature on framing and complexity effects in the CE literature and relates this to spatial choices for environmental valuation. Section 3 outlines our methodological and modelling approach. The case study is described in Section 4, followed by the results in Section 5 and conclusions in Section 6.

2. Framing and complexity

The question of how people take substitute sites into account is highly relevant for SP studies. For the reliable estimation of substitution effects between different sites, the choice set specification in CEs is hence an important step in the study design. The choice set specification can affect parameter estimates (Pellegrini et al., 1997; DeShazo and Fermo, 2002) and resulting economic welfare estimates (Parsons and Hauber, 1998). Choice set specification is particularly important for spatial choice studies for environmental valuation, as the number of relevant alternative locations in the study area can be large (Parsons and Kealy, 1992; Cummings et al., 1994).

The assumption that respondents have consistent preferences has often been violated in empirical studies. The lack of substitute effects on WTP in these studies has been attributed to survey-instrument bias or framing effects, including insensitivity-to-scope, part-whole and embedding bias (e.g., Carson and Mitchell, 1995; Carson and Groves, 2007). Framing effects occur when providing information in different formats leads to different choice outcomes (Kahneman, 2002). Framing is expected to affect choices especially when individuals are unfamiliar with the good or market setting in which the good is provided (Scott, 2002). Environmental goods and services in SP research are a typical example. The selection of alternatives in the CE choice set involves a framing decision, as it determines how a study site is presented in relation to its substitutes. In other cases, apparent violations of rationality assumptions have been related to anomalies in decision-making, arising when rationality assumptions are not met, for instance, due to preference uncertainty (Braga and Starmer, 2005; Campbell et al., 2008).

² We do not apply spatial econometric methods in this paper. We use the term "spatial choice studies" to refer to studies in which choices are based on goods and services provided at different geographic locations, where this location, the distance between the alternative locations and distance between locations and respondents are expected to play an important role in the choice behaviour.

Anomalies may result from choice task complexity, which is determined by the number of choice tasks, alternatives within a task, attributes within an alternative, levels within an attribute, and the range of the attribute levels (Caussade et al., 2005). Complex choice settings with a large number of alternatives in the choice task may impose a high cognitive burden on respondents. If the cognitive burden (related to the amount of information that has to be processed to make a choice), exceeds the cognitive ability of the respondent, choice task complexity may prohibit the respondent from choosing the most preferred alternative (Caussade et al., 2005). Swait and Adamowicz (2001) show that complex choice scenarios result in choices that are less consistent with rationality assumptions than simple scenarios. Swait and Adamowicz (1996) find that as the choice tasks become more difficult, people are more likely to choose the opt-out and focus more on a product's brand than on its attributes. This suggests that respondents may apply simplifying decision-rules by choosing an alternative based on their ranking of one of its attributes that they consider most important (Alpizar et al., 2001). If task complexity becomes too high, respondents may apply simplifying information processing and decision-making strategies (e.g., Payne et al., 1993; Puckett and Hensher, 2008), which violate the assumption that all alternatives are evaluated simultaneously based on a trade-off of all characteristics. Some studies have found that when respondents are asked to answer a series of questions, they initially learn about their preferences and the CE setting, and familiarise themselves with the choice tasks (Swait and Adamowicz, 2001; Holmes and Boyle, 2005). After this learning process, choices have been found to become stable. The final choices of the choice task series may show inconsistencies again as respondents become fatigued (e.g., Kontoleon and Yabe, 2003), but the empirical results are mixed (Ohler et al., 2000; Savage and Waldman, 2008). It is, however, difficult to separate learning effects from fatigue. At most, their net effect can be analysed, since they may offset one another.

The effect of the number of alternatives on choice stability has been analysed in cross-section studies (i.e. between-respondents). Arentze et al. (2003) did not find empirical evidence supporting the hypothesis that a larger number of alternatives leads to higher error variance or different attribute weights. DeShazo and Fermo (2002) and Caussade et al. (2005) find a quadratic relationship between the number of alternatives and the model variance. Their between-respondent tests suggest that the effect of the number of alternatives on the error variance follows an inverted U-shaped pattern. The model variance initially decreases suggesting that preferences and choices become better-defined as more alternatives are added to the choice set. However, the variance increases when choice sets include more than three to four alternatives, indicating lower choice precision due to a higher complexity. Fotheringham and O'Kelly (1989) suggest that choice sets of six or more alternatives will lead to non-compensatory choices across sites in spatial choice studies. These results suggest that in order to obtain reliable WTP estimates, a choice set size of three to four alternatives would be optimal in CEs.

Besides choice complexity, another important criterion in choice set specification is relevance. The 'relevant choice set' or 'consideration set' includes the substitutes that are relevant to the choice process of the respondent and therefore should be included in the design and analysis. The relevant set of substitutes varies across goods and sampled populations (Boyle and Bergstrom, 2001). It is difficult to identify the 'relevant choice set' without *a priori* knowledge about respondent perception. Some studies include all sites known or considered to be relevant (Peters et al., 1995; Haab and Hicks, 1997; Parsons and Hauber, 1998). Mostly, however, researchers have used different decision rules to determine the relevant alternatives. Recreational site choice studies, for example, have selected all sites of a similar type, all outdoor recreation sites in an area, or all recreation and leisure options (e.g., Jones et al., 2010). Alternatively, the identification of relevant alternatives can be modelled as an endogenous part of the choice process (e.g., Swait and Ben-Akiva, 1987; Ben-Akiva and Boccara, 1995), based on Manski's (1977) two-stage choice process, in which the probability of choosing an alternative is conditional on the probability that the alternative is considered and included in the individual's relevant choice set.

The main challenge in the design of the experimental choice set and selection of alternatives is to find the right balance between the relevance and complexity of the choice task in view of the cognitive capabilities of the respondents.³ The analyst has to decide which alternatives of the 'relevant choice set' of the respondent to include in the 'experimental choice set' of the CE. A larger experimental choice set may increase choice complexity and affect the reliability of the results. Including many alternatives may serve to assure that respondents consider all available options and overcome the problems of single site studies. Hensher et al. (2005b) argue that the increase in distance may in fact facilitate choices, as the alternatives differ more in "attribute space" and are hence easier to distinguish. If all included alternatives and attributes are relevant, respondents may be able to handle more complex tasks (Rose and Hensher, 2004).

3. Experimental design, hypotheses and modelling approach

3.1. Experimental design and hypotheses

In our analysis, we use data from a labelled, web-based CE focusing on the non-market benefits that households attach to improved ecological quality in eleven lakes at different distances from the respondents' homes.⁴ The CE is decomposed into three rounds with choice sets of different sizes. In the first round of the CE, consisting of three choice tasks, respondents are asked to choose among ecological improvement scenarios at the four lakes closest to their residential location with

³ There are also practical limitations to the number of alternatives and attributes that can be included in the survey design of a CE, given the statistical efficiency of the experimental design and the potential sample size.

⁴ In the marketing literature, the term 'branded' is commonly used instead of 'labelled'. Here, the names of the lakes will be associated with the characteristics of the lakes that are not captured by the attributes.

associated price increases. In the second round of the CE, consisting of four choice tasks, the choice sets are expanded to include seven lakes randomly drawn from the full set of eleven lakes based on the experimental design. The alternatives added to the choice sets are quality changes taking place at lakes located further away, which are therefore also likely to be less familiar to respondents, thereby possibly increasing the complexity of the choices (or not as argued by Hensher et al. (2005b)). In each choice task, different combinations of seven of the eleven lakes with associated quality and price increases are presented as eligible alternatives. In the third round, the respondents are presented with two final choice tasks focusing on ecological improvement scenarios at the four nearest lakes to their home again. The ecological improvement scenarios and associated price increases at these four lakes are not the same as in the first part (which would have allowed us to test choice consistency) since we are primarily interested here in testing preference and WTP stability. The choice sets in the first and third round of the CE will be referred to as “small choice sets” and those in the second round as “large choice sets”.

With this study design, three sets of hypotheses are formulated to test the effect of the geographical framing of the choice set on preference stability and WTP. First, a comparison is made between the choices among alternatives in small and large choice sets and resulting WTP values for ecological improvements for the same lakes. This comparison provides a test of the effect of the geographical scale of choice sets on WTP, with embedded tests of learning and complexity effects.⁵ The following hypotheses are tested:

H_0^1 : WTP for water quality changes at lake i is independent of the geographical scale of the choice sets in which it is embedded: WTP_i (small choice sets) = WTP_i (large choice sets).

H_0^{1a} : Preference parameters do not differ between small and large choice sets.

H_0^{1b} : Scale parameters do not differ between small and large choice sets.

Rejection of hypothesis H_0^1 would imply that choice sets including more alternatives from a larger geographical scale lead to different WTP results. Different model results for choices based on large and small choice sets would call into question the validity and reliability of parameter estimates. Under assumptions of well-formed choices, it is expected that the marginal value of a quality change at one of the lakes does not change after the inclusion of other improvement scenarios at other lakes in the choice sets. However, changes in preferences for the attributes may be observed if the complexity of the choice task in the large choice sets, where respondents have to evaluate additional substitutes, becomes too high or when underlying preferences for the alternatives change. Therefore, we examine whether adding alternatives increases the complexity of the choice task and consequently leads to instability of choices and preferences, measured as changes in coefficients of the attributes (H_0^{1a}).

Another effect of task complexity and increasing the choice sets can manifest itself in changes in the error variance (Mazzotta and Opaluch, 1995), leading to hypothesis H_0^{1b} . Given that the scale parameter is inversely related to the variance of the error term, scale increases indicate that the variance in the model decreases (Louviere et al., 2000; Swait and Adamowicz, 2001). This suggests that choices become more deterministic. On the other hand, more difficult choices are associated with lower scale parameters and higher overall model variance. This is expected to be the case for the large choice sets in this case application. Learning is reflected in a decrease in the variance relative to the initial choices in a series of choice tasks.

The second set of hypotheses addresses potential differences between choices among alternatives in the first and second round with small choice sets. This provides a test to see if respondents change their WTP after consideration of additional locations included in the large choice set. Embedding the small set in the large set with more substitutes requires respondents to actively evaluate these additional substitutes and the spatial setting of the choice task, rather than simply taking into account any information provided in the explanatory text about the quality and location of these substitutes. Thinking about additional substitutes may furthermore help respondents to learn and refine their preferences for the alternatives in the second round with the small choice sets. The hypotheses to be tested are:

H_0^2 : WTP for water quality changes does not change after evaluating large choice sets: WTP (1st round with small choice sets) = WTP (2nd round with small choice sets).

H_0^{2a} : Preference parameters do not differ between the 1st and 2nd round with small choice sets.

H_0^{2b} : Scale parameters do not differ between the 1st and 2nd round with small choice sets.

Differences in WTP values between the first round of the CE with small choice sets and the last round of the CE with small choice sets, after respondents have considered several intervening large choice sets, may indicate that the preferences of respondents regarding ecological quality improvements in the lakes in small choice sets have changed after consideration of the substitutes in large choice sets. If attribute parameters are not significantly different between the first and second round with small choice sets, as tested in H_0^{2a} , the conclusion would be that respondents have well-formed preferences. If the variance is not significantly different either (H_0^{2b}), this would suggest that there is no net effect of learning and fatigue effects present. It would imply that respondents do not find the choice tasks in the second round with small choice sets

⁵ As a consequence of the study design, learning and complexity effects are confounded when comparing the choices in the first round with small choice sets and the choices with large choice sets.

more demanding (expected to increase variance), do not learn from choices made in large choice sets when choosing again among small choice sets (expected to reduce variance) and do not suffer from fatigue after considering the choice tasks in the large choice sets (expected to increase variance), or that these effects cancel out.

3.2. Modelling approach

The tests will be based on the estimation results of mixed logit models (Brownstone and Train, 1998). Random parameter logit (RPL) models assume that the coefficients follow a specific underlying distribution instead of being fixed, accommodating preference heterogeneity in the population. Thereby, mixed logit models avoid the IIA assumption. Additionally, common random parameters with zero mean, i.e. error-components, can be included in the utility function specification of those alternatives that are likely to be correlated, to allow for correlation in the unobserved utility between alternatives (Brownstone and Train, 1998). Error-components can be used when comparing less familiar or hypothetical alternatives with better known ones or the opt-out (Scarpa et al., 2005). Similarly, they can be used to control for scale-differences between datasets in studies combining different datasets, such as SP and RP data (e.g., Brownstone et al., 2000; Hensher et al., 2005a). Mixed logit models are also used to control for the panel structure of the data (Scarpa et al., 2005).

The large choice sets in this CE include alternatives that are relatively far away from the respondents' residences. The travel distance via the existing road network between respondents and alternative lakes is expected to play a pivotal role in choice behaviour. In SP studies, distance has been put forward as a validity indicator of site valuation studies and used as a proxy for the travel costs associated with visiting the lake (Bateman et al., 2011). Distance is hence expected to have a negative effect, which reflects that WTP decreases with distance. This pattern is known as distance-decay (Sutherland and Walsh, 1985). In this study, we control for heterogeneity in distance-decay across respondents. First, to capture unexplained heterogeneity in distance decay, the coefficient on distance is included as a random parameter in the models. Heterogeneity may, for instance, be caused by individual differences in experience or familiarity with the sites or in the perceived cost of travelling. Systematic differences have been found between visitors (users) and non-visitors of sites, where distance-decay effects are expected to be lower or close to zero for respondents with mainly non-use values, because travel costs are theoretically expected to have a smaller effect on their decisions compared with users. Empirical studies have, however, found significant distance-decay in non-use values, explained by distance-related salience and available information (Schaafsma, 2010). We test the effects of use(r) and non-use(r) related preferences on distance decay in WTP for water quality changes in our analysis.⁶

In the basic model, the utility function of individual i for lake n is specified as a function of the price of a lake, P , the characteristics of the lake, X_{nt} , as well as the alternative-specific constant, α_n , where t are the different choice tasks ($t=1, \dots, 9$):

$$U_{int} = \alpha_n + \beta^p P_t + \beta_n^x X_{nt} + \varepsilon_{int}$$

In this experiment, the attribute in X_{nt} represents the changes in environmental goods and services provided by the different sites, and has two levels related to different quality levels (see Section 4). We then allow the coefficient on X_{nt} to depend upon the distance from the respondent to the site, D_{in} :

$$U_{int} = \alpha_n + \beta^p P_t + (\beta_n^x + \beta_{in}^d D_{in}) X_{nt} + \varepsilon_{int}$$

Then, the coefficient on $D_{in} X_{nt}$ is allowed to vary systematically with a vector of respondent characteristics, Y_i , which reflect use and non-use related preferences for site n , producing:⁷

$$U_{int} = \alpha_n + \beta^p P_t + (\beta_n^x + (\beta_{in}^d + \beta^y Y_i) D_{in}) X_{nt} + \varepsilon_{int}$$

Finally, the error term is generalised to include a zero-mean error-component that is shared by all of the hypothetical alternatives, given by $\lambda_i^h d_{in}$, where d_{in} takes the value 1 for all of the hypothetical alternatives n and zero for the status-quo option:

$$\text{Model I: } U_{int} = \alpha_n + \beta^p P_t + (\beta_n^x + (\beta_{in}^d + \beta^y Y_i) D_{in}) X_{nt} + \lambda_i^h d_{in} + \varepsilon_{int}$$

The parameter λ_{in} is assumed to have a normal distribution $N[0, \sigma]$. The term ε_{int} reflects the error term of the model, which is assumed to have an i.i.d. extreme value type 1 distribution. Our final first model (Model I) is hence specified as a mixed logit model that accounts for the panel structure of the data and for the difference in error terms between the hypothetical alternatives and the status quo. For estimation purposes, the model is expressed as a linear combination of coefficients and variables:

$$U_{int} = \alpha_n + \beta^p P_t + \beta_n^x X_{nt} + \beta_{in}^d (X_{nt} D_{in}) + \beta^y (X_{nt} D_{in} Y_i) + \lambda_i^h d_{in} + \varepsilon_{int}$$

Model II is also a mixed logit model, but estimated based on the pooled dataset of the large and small choice sets. To examine differences in preferences for environmental changes in the lakes between small and large choice sets and

⁶ See Ferrini et al. (2008) for an alternative approach to separate use and non-use values and spatial complexity in environmental valuation studies by combining travel cost and choice experiment data.

⁷ Our interest in revealing heterogeneity in distance-decay related to the WTP for quality changes, and the effect of use and non-use motivations involved is the main reason to include the combination of the attributes in X and the respondent characteristics D and Y .

between the first and second round with small choice sets, the quality attribute levels are allowed to vary with two dummy variables, L and Sb . Dummy variable L takes the value 1 for large choice sets ($t=4,5,6,7$), where respondents choose among water quality improvement scenarios at seven out of the eleven lakes in the full choice set, whilst Sb equals one for all choice occasions in the second round with small choice sets ($t=8,9$). Furthermore, to capture possible scale-differences, the error term of the model is further generalised to include a zero-mean error-component shared by all hypothetical alternatives in large choice sets, given by $\lambda_{in}^L L_{in}$, and another zero-mean error-component shared by all hypothetical alternatives in the second round with small choice sets, $\lambda_{in}^{Sb} Sb_{in}$.⁸ As before, the λ_{in} parameters are assumed to have a normal distribution $N[0, \sigma]$. This produces:

$$\text{Model II: } U_{int} = \alpha_n + \beta^p P_t + (\beta_n^x + \beta^L L + \beta^{Sb} Sb + (\beta_{in}^d + \beta^y Y_i) D_{in}) X_{nt} + \lambda_{in}^h d_{in} + \lambda_{in}^L L_{in} + \lambda_{in}^{Sb} Sb_{in} + \varepsilon_{int}$$

In the case study, the attribute parameters δ_n^x are expected to be positive, because X represents quality improvements that are expected to increase utility. The coefficient β^p , related to price P (tax increase), is expected to be negative. It is included as a fixed parameter to reflect constant marginal utility of income in order to accommodate welfare comparison across respondents, and to avoid non-negative individual specific parameter estimates that can arise when the standard error of the random parameter is very large.

3.3. Hypotheses tests

As a first test of hypothesis H_0^1 , Model I is estimated separately for small and large choice sets. The resulting WTP estimates and their confidence intervals are compared. Overlapping confidence intervals imply that the WTP estimates are not significantly different. A second test for hypotheses H_0^{1a} , H_0^{1b} , H_0^{2a} and H_0^{2b} is performed using the Swait and Louviere (SL) test for differences in the attribute and scale parameters (Swait and Louviere, 1993). This test can be applied to control for changes in preferences over a sequence of choices (e.g., Carlsson et al., 2010; Brouwer et al., 2010), when combining data from RP and SP studies (e.g., Adamowicz et al., 1994), or to compare differences in parameters across different populations (e.g., Colombo et al., 2007).

The SL test involves a number of sequential steps. First, a test for changes in beta parameters (β) is performed by allowing for varying scale parameters between two (sets of) choice tasks. If the null hypothesis of beta parameter equality is rejected, there may be differences between the two (sets of) choice tasks, but it will be impossible to attribute these to differences in either the beta parameters alone or the beta and scale parameters, because the scale and beta parameters are confounded (Louviere et al., 2003). If the null hypothesis cannot be rejected, a second test for scale parameter equality is performed.

For the first test, we estimate two separate models for the (sets of) single choice tasks, which gives to sets of scale and beta parameter estimates and the log-likelihoods. Then, a pooled model based on the two (sets of) choice tasks together is estimated, which imposes beta parameter equality. For this pooled model, a grid search for the scale parameter is performed to optimise the log-likelihood under different relative scale adjustments for the second choice task (or set of choice tasks). Using the results of the pooled model with the best model fit, a chi-square (Likelihood Ratio) test is performed to see if restricting the beta parameters to be equal for the two (sets of) choice tasks results in a significantly different model fit.⁹ If the null-hypothesis of equal beta parameters is not rejected, the second test can be performed for differences in scale parameters. A pooled model is estimated where beta and scale parameters are restricted to be identical. Another LR test (d.f. equal to 1) is then applied to compare the log-likelihood of this model with the pooled model with a different scale parameter.

Alternatively, Brownstone et al. (2000) and Hensher et al. (2005b) suggest that the distinction between differences in the scale and beta coefficients can be made using error-components in a mixed logit model specification. Therefore, the pooled Model II is estimated. Significant estimates for the interaction terms of the quality attributes and choice set size (β^L and β^{Sb}) indicate preference changes in the quality attribute. Statistically significant parameter estimates for the coefficients that contribute to the error-components λ_{in}^{Sb} and λ_{in}^L in Model II indicate that the variance of the observations in these subsets of choices is different from the variance for the initial three choice sets consisting of four alternatives.

As an alternative to the pooled Model II, a more general specification would be to estimate a model similar to Model I for the three parts of the CE separately and then combine these in an all-inclusive log-likelihood function. Treating the three parts as independent sub-samples would ignore, however, the correlation between choices of the same respondent in the estimation of the individual-specific random parameters and the error-term, and is therefore deemed inappropriate for our application.

⁸ The error-components related to L and Sb take the value zero for the opt-out for identification, because they are included in the model as individual specific random parameters and, being a dummy variable, must have a baseline for each individual choice.

⁹ The LR-test is as follows: $-2(LL(A\&B)) - (LL(A) + LL(B))$ with d.f. $(K+1)$. A and B are the different choice tasks. K is the number of parameter restrictions (and therefore equal to the number of parameters in the models). The additional degree of freedom is because the scale parameter is allowed to vary under the alternative hypothesis (Swait and Louviere, 1993).

4. Case study design

The data used in this paper were collected using a web-based CE, which aimed to estimate the WTP for environmental changes in a lake district in the central-west region of the Netherlands, lying amidst the cities of Amsterdam, Utrecht and The Hague. The area is known as the Green Heart, a relatively thinly populated area with a rural character. The eleven lakes included in the experiment are of reasonable size (> 200 ha) and provide a range of water recreation possibilities, including boating, fishing, bathing and nature watching. The area provides a habitat for a number of IUCN red list (globally threatened) bird species. Although they provide similar recreational possibilities, the lakes lie up to 60 km apart. Fig. 1 presents a map of the study area and depicts the eleven lakes.

The lakes suffer from eutrophication due to excessive nitrogen and phosphorus loads from the surrounding agricultural lands. As a result, most lakes are of moderate quality, with the exception of the Naardermeer, which has a good ecological quality. In this study, the good under valuation consists of the use and non-use values of ecological quality improvements at the eleven lakes following the implementation of the European Water Framework Directive (WFD). The objective of this directive is to achieve 'good ecological status' by 2015. This status is associated with use as well as non-use values, which justifies the use of SP methods (Brouwer, 2008).

Following the policy objectives, three different ecological equilibrium states were formulated as possible future scenarios for the lakes: the current status, the 'good ecological status' and an intermediate ecological state. To convey these states in terms of changes in ecosystem service provision of the sites, the quality levels were specified in terms of the number and diversity of fish and bird species, water clarity and corresponding recreational opportunities. The nature watching and bathing amenities are expected to increase in the intermediate and good ecological state, but require limits on recreational boating in order to reach a better ecological state compared to the current state. Two important issues have to be considered when including water quality as a single, comprehensive attribute in the choice experiment. First, the levels of recreation are collinear to some degree, either because they depend on the ecological state or because achieving the ecological objectives require certain limitations on recreational activities. Rather than generating a CE design with individual attributes for different recreational activities and deleting implausible combinations, we combined the quality levels and associated recreational amenities into a single attribute. Although this prohibits eliciting separate values for these recreational

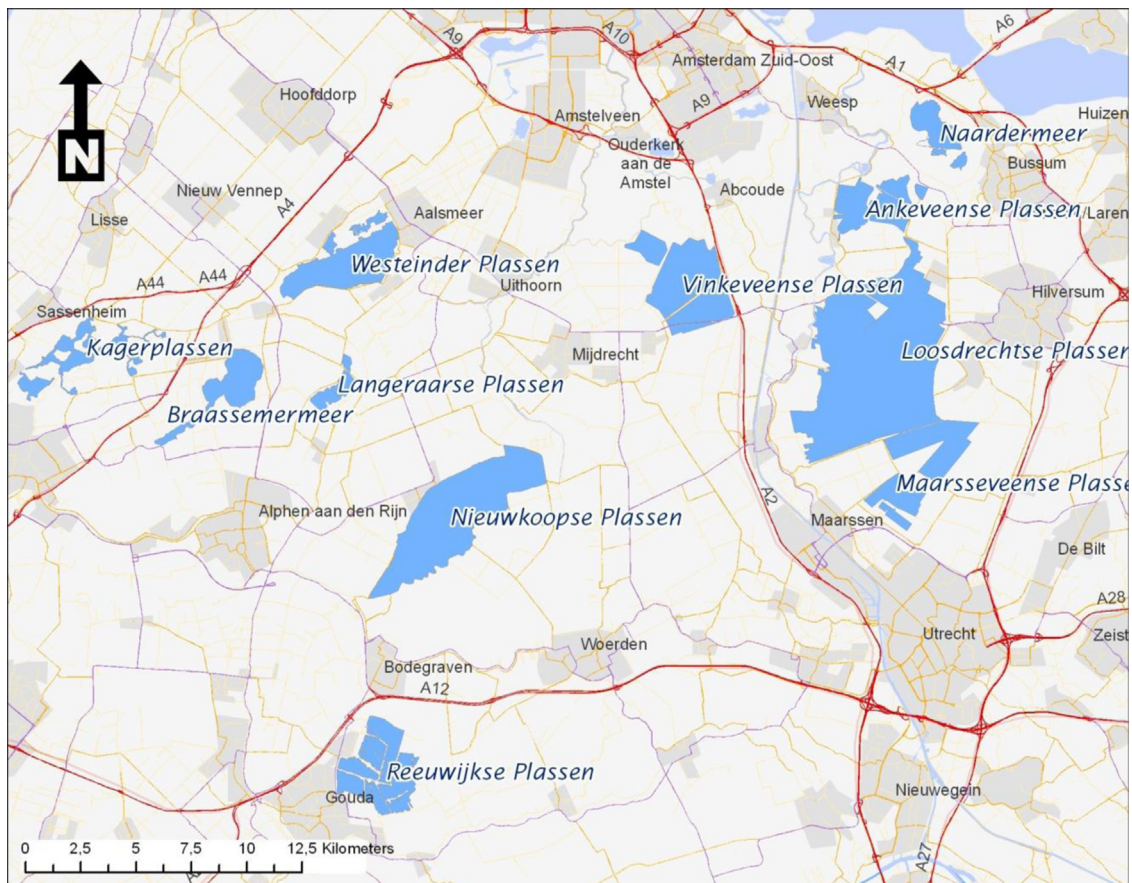
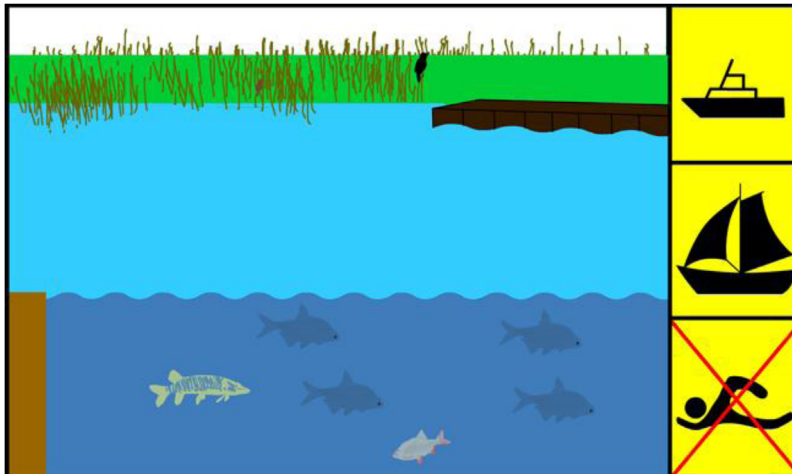
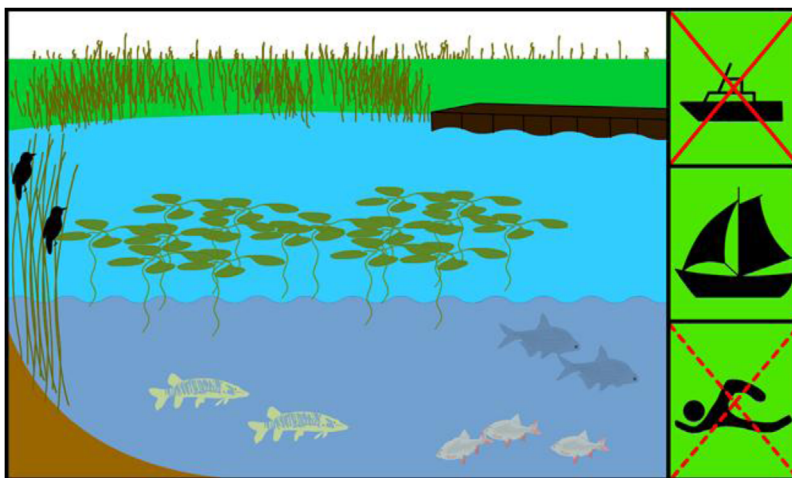


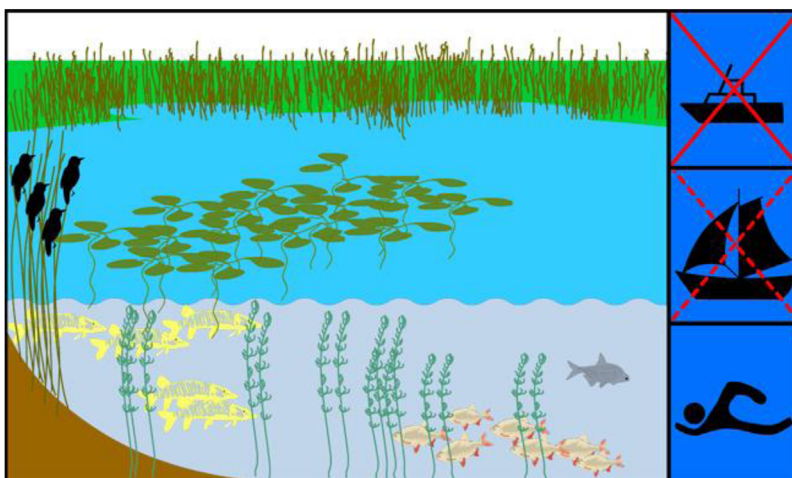
Fig. 1. Map of the case study area.



At the **YELLOW** level, the water is turbid, and you can see less than a meter deep. There are few birds, especially few endangered bird species. There are many breams, but few other fish species, such as pike. Reed grows along some of the banks. Bathing is often prohibited due to toxic algae blooms. Sailing and motorised boating is allowed and there are many piers.



At the **GREEN** level, the water is rather clear and visibility is about one meter. There are some breams and pikes. A small number of endangered bird species are present. There are some water plants and reed is found along the banks. Due to toxic algae, bathing is prohibited a couple of times each summer. Motorised boating is prohibited, but sailing is possible and piers are available.



At the **BLUE** level the water is very clear. There are many fish species, primarily pike. There are also various protected bird species present, such as the reed warbler. There are many water plants and thick reed areas along most of the banks. Swimming is possible during the entire summer. There are more shallow areas, in which sailing is not possible. Motorised boating is prohibited.

Fig. 2. Ecological quality descriptions used in the choice experiment.

activities, pre-tests indicated that a design with a single attribute comprising the activities was preferred over a choice experiment with different attributes and combinations of levels that respondents deemed unrealistic. Furthermore, multiple quality-driven attributes would have complicated the design and prohibited displaying the lakes and attributes on a map of the study area, which was considered to be an important design aspect to reveal spatial preferences. Similar water quality indicators and ladders have been used in other studies, e.g., [Bateman et al. \(2011\)](#). Secondly, due to the trade-off between recreational and conservation activities required to achieve higher ecological quality levels, the optimal ecological status is not necessarily associated with the highest WTP. The preference ordering of the two improvement levels is expected to be subject to taste heterogeneity, depending on the type of recreation or amenity valued most by individual respondents.

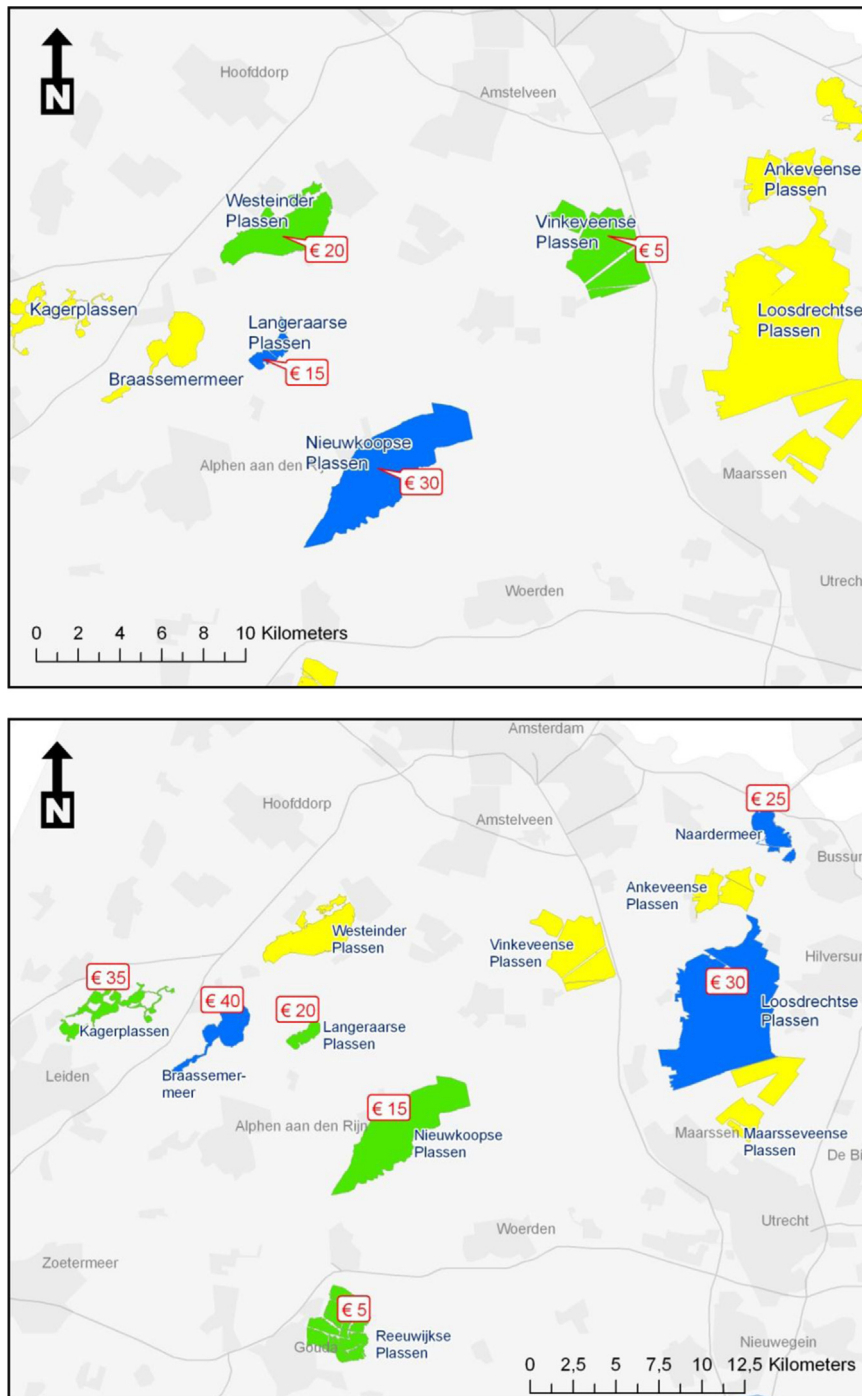


Fig. 3. Example choice cards of small (above) and large (below) choice sets.

As a result, the study design does not lend itself directly to a scope test of WTP for water quality improvements (Banerjee and Murphy, 2005).

The description of the three ecological states was presented in illustrations with pictograms reflecting the possibilities for motorised boating, sailing and bathing (see Fig. 2). The three levels were colour-coded: yellow for the current state, green for the intermediate, and blue for the optimal ecological state. These colours were thoroughly pretested and used in the design of the choice cards to depict the future quality levels at the different locations along with the respective prices.

A labelled CE was used. Treating the different locations as separate alternatives allows controlling for changes in the characteristics of sites on the WTP for alternatives. Respondents were asked to choose between the sites under various scenarios of ecological quality improvements against an additional payment. Each choice card depicted the goods under valuation at the site where they are provided in relation to the location of its substitutes. An example of a choice card is shown in Fig. 3. In the survey the illustrations in Fig. 2 were shown alongside each choice card to ensure that respondent clearly understood the different quality levels. Hence, the design of the CE not only includes the presence of substitutes, but also their location, quality and price. Making these characteristics of the choice context explicit in the experimental design is thereby expected to overcome the limited focus and corresponding disadvantages of single site studies.

Price levels ranged from 5 to 40 Euro per household per year. The payment vehicle was an increase in the annual water tax that every household has to pay. Previous studies concerning water valuation in the Netherlands found that this payment vehicle is preferred by respondents (Brouwer, 2008). Each choice task included an opt-out option, in which quality remained at the current level and taxes did not increase. The accompanying survey text explained that the lakes that were not chosen by the respondent would remain at their current quality level and the tax increase would only be spent on the ecological quality improvement of the chosen alternative.

Each respondent was given nine choice tasks in total. For small choice sets, the same D-efficiency design with five fixed blocks was used. Only the labels were changed so that each respondent was offered to choose among the four nearest lakes. For large choice sets, the choice tasks were based on a D-efficiency design with 30 fixed blocks. The order of the choice tasks was not randomised: all respondents received three choice tasks with small sets, then four with large choice sets, and then two again with small choice sets. The survey was finalised after various rounds of pre-testing: through a focus group discussion, face-to-face interviews and an online pre-test survey. The questionnaire consisted of three main parts with questions about: (a) public perception of lake quality and recreational opportunities, (b) the CE, and (c) socio-demographic characteristics. The data were collected in March 2009 using an online panel. We aimed to select respondents who lived within a 40 km range (based on travel distances via the existing road network) of all lakes in each group of four (small choice sets). To this end, we estimated the road distance from the centroids of 4-digit postal code areas to various entry points along the lake boundaries using GIS, and recruited respondents from postal areas located within the relevant range. Two reminders were sent to obtain sufficient responses, ensure representativeness of the sample for the population living in the area and obtain sufficient variation in the spatial distribution of the sample.

5. Choice experiment results

5.1. General survey results

The total number of respondents completing the survey is 889. It took respondents on average 15 min to complete the survey. The modal education level of the sample is higher vocational training, 15 percent of the respondents completed a middle vocational training and 17 percent hold a university degree, similar to regional socio-economic statistics (Ruimtemonitor, 2008). The modal household income is between €31,500 and €63,000 per year, in line with the provincial statistics (CPB, 2010). These statistics suggest that, based on observable key indicator variables, the sample is representative for the population in the sampling area.

Eight percent of the sample never visit open fresh waters and another 11 percent have not visited any fresh water bodies over the last 12 months.¹⁰ The most popular recreational activities at open waters are walking, running and cycling, followed by bird and other wildlife watching, and bathing. A third of the respondents visit lakes for sailing, surfing, canoeing or rowing, and only a quarter engages in motorised boating and waterskiing. Boaters will face limitations of their recreational activities if the higher ecological quality levels are achieved. Although two-thirds of the respondents consider the current water and nature quality conditions of the lakes to be good in general, an equally large proportion of the sample (88 percent) finds further improvements important. Twenty-four percent of the respondents want to contribute to better water quality because they visit the lakes and 33 percent want to improve the conditions for flora and fauna. Nine percent of all respondents choose the opt-out because of protest reasons, mostly because they believe the water managers or government should finance the plans with the current budgets. Respondents were asked on which factor(s) they mainly focused when making choices. The results show that 44 percent considered the location as one of the main arguments when making choices in the CE. The distance to the lakes is one of the main determinants for 23 percent of the respondents. Another 22

¹⁰ Here, 'open fresh water bodies' refer to any outdoor water body, except the sea, that people may use for recreation, including lakes, rivers, ponds, larger ditches, canals, etc. Use values here relate to both contact and non-contact (shoreline) activities.

Table 1a

Results of Models I and II for the pooled database.

Lakes	Model I: small choice sets (choices $t=1,2,3,8,9$)	Model I: large choice sets (choices $t=4,5,6,7$)	Model II: pooled (choices $t=1,...,9$)
Alternative-specific constants			
Kagerplassen (KA)	0.075 (0.521)	0.473** (2.492)	0.338*** (3.801)
Braassemmermeer (BR)	0.262*** (2.343)	0.405** (2.206)	0.332*** (4.360)
Westeinder Plassen (WE)	0.267** (2.342)	0.378** (2.089)	0.377*** (4.688)
Nieuwkoopse Plassen (NK)	1.426*** (13.387)	1.245*** (7.105)	1.309*** (18.400)
Reeuwijkse Plassen (RW)	1.520*** (10.252)	1.078*** (5.996)	1.305*** (14.076)
Vinkeveense Plassen (VV)	0.773*** (6.178)	1.091*** (6.082)	0.949*** (12.544)
Naardermeer (NM)	1.603*** (9.462)	1.288*** (6.716)	1.424*** (16.478)
Ankeveense Plassen (AV)	0.773*** (6.178)	0.643*** (3.333)	0.909*** (10.407)
Loosdrechtse Plassen (LD)	1.899*** (11.772)	1.439*** (8.041)	1.639*** (20.817)
Maarsseveense Plassen (MV)	1.048*** (6.101)	0.726*** (3.861)	0.852*** (9.684)
Random parameters (normal distr.)			
Quality \times Distance (ρ_{in}^d)	-2.659*** (-22.142)	-1.878*** (-18.665)	-2.086*** (-29.946)
Standard deviation	1.540*** (15.055)	0.993*** (8.313)	1.142*** (18.834)
Large set (zero mean): standard deviation (λ_{in}^L)			2.038*** (11.913)
Small set second part (zero mean): standard deviation (λ_{in}^{Sb})			1.040*** (4.719)
Non-random parameters			
Price	-0.383*** (-18.998)	-0.040*** (-16.739)	-0.038*** (-32.375)
Intermediate ecological quality level (green)	7.122*** (18.793)	5.031*** (10.650)	6.121*** (19.938)
Optimal ecological quality level (blue)	7.288*** (19.290)	5.067*** (10.706)	6.298*** (20.753)
Quality \times Distance \times user (dummy)	0.482*** (20.561)	0.438*** (16.662)	0.430*** (33.192)
Quality \times Distance \times non-use reason (dummy)	0.869*** (4.878)	0.597*** (4.903)	0.580*** (5.207)
Intermediate ecological quality level \times Large set (dummy)			-0.098 (-0.567)
Optimal ecological quality level \times Large set (dummy)			-0.225 (-1.287)
Intermediate ecological quality level \times Second part small set (dummy)			-0.098 (-0.604)
Optimal ecological quality level \times Second part small set (dummy)			-0.179 (-1.091)
Error-component			
All lakes (λ_{in}^h)	5.210*** (13.387)	3.895*** (42.950)	5.648*** (20.407)
Model statistics			
No. observations: choices (respondents)	3880 (776)	3184 (796)	6831 (759)
Log-likelihood	-3919	-4206	-7556

Notes: *T*-values between brackets. "Quality" refers to both ecological quality levels when included in the model in combination with other variables and no quality level-specific parameters are estimated. The variable "Distance" is included in the model using a natural log transformation: $\ln(\text{km}+1)$.

*Significance of the estimates at 10%.

** Significance of the estimates at 5%.

*** Significance of the estimates at 1%.

percent of the respondents base their choices on the combination of price, quality and location. Fifteen percent of the respondents mention quality and the same percentage is found for the price of the alternatives.

5.2. Expanding the geographical scale of the choice set

The results of the CE were analysed using NLOGIT 4.0 and the estimates are presented in Table 1a. The results indicate that the sample population attaches a positive WTP to environmental improvement scenarios. The two ecological quality levels (intermediate and optimal) are included as dummy variables in the model and their coefficients are interpreted as changes in marginal utility compared to the status quo. The coefficients of the two levels of the quality attribute are positive, but they are not significantly different from each other, which indicates that respondents do not attach a significantly higher value to the optimal ecological quality level compared to the intermediate ecological quality level.¹¹ An immediate policy implication of this finding is that the public benefits related to the optimal ecological quality level, reflecting the WFD objective of 'good ecological status', are not significantly higher than intermediate levels of water quality. This result may be due to the fact that the improvements in recreation and nature amenities related to the 'good ecological status' insufficiently compensate for the loss of recreational value due to the restrictions imposed on recreational boating required to achieve this optimal ecological quality level, compared with the intermediate level of ecological quality.¹² Since achieving the good

¹¹ The final models nevertheless include individual variables for the two quality levels, because a model with two separate coefficients results in a better model fit than if a single coefficient is estimated for the sum of the mutually exclusive quality levels (thereby restricting their coefficients to be equal). In addition, a model specification with two separate quality coefficients demonstrates more clearly the policy implication of this finding.

¹² We tested if there were significant differences in WTP for the (blue) optimal ecological quality level between people who prefer boating (and may not have higher WTP for the optimal ecological quality level than for the intermediate ecological quality level), and other recreationists, but no statistically significant difference was found. Bathers did not have higher WTP for the optimal level either.

Table 1b
Cholesky matrix of pooled Model II.

	Quality × Distance	Large choice sets—error component	Second round small choice set—error component
Quality × Distance	1.142*** (18.834)		
Large choice sets—error component	−0.191*** (0.903)	2.029*** (11.939)	
Second round small choice set—error component	0.542*** (2.633)	−0.778 (4.437)	0.426 (1.120)

Notes: T-values between brackets.

***Significance of the estimates at 10%.

ecological status is likely to require more investments, the important policy question rises if achieving this optimal ecological status will result in higher net public benefits than realising less ambitious ecological objectives.

As expected, WTP declines as respondents live further away from the sites, indicated by the negative parameter for the quality-distance interaction.¹³ WTP is higher, and distance decay is lower, for respondents who have visited the sites providing the environmental goods and services. The distance decay in the WTP for non-users most probably reflects the presence of option values, i.e. the opportunity to use the sites in the future under improved ecological conditions. Furthermore, WTP is higher for respondents whose main reason for contributing to better ecological quality is motivated by non-use reasons, as indicated by the positive coefficient for the interaction term between the dummy variable taking the value 1 for respondents with non-use related reasons for WTP and the quality-distance interaction term.¹⁴

5.2.1. Test 1: separate models

Of main interest are the results for the choice set effects. The hypotheses were tested in three different ways. As a first test of hypothesis H_0^1 , the WTP results of Model I estimated separately for small and large choice sets are compared, and the overlap of the confidence intervals of the WTP estimates is examined. By taking the ratio of parameter estimates of the quality and distance attributes and the price, any difference in scale between the subsets cancels out. The WTP results are presented in Table 2.

The comparison of the WTP implications of the models for large and small choice sets suggests that the WTP for quality changes is lower in large choice sets. However, the confidence intervals of the WTP values of the two choice set specifications overlap after 3 km from the site as a result of the weaker distance-decay effect found for the large set. This is because the WTP for quality changes is also dependent on the distance effect, which is included as an interaction term with the quality levels. The graph included in Table 2 shows that the WTP values of small and large choice sets have significantly different values at zero distance, but converge as distance increases. The dotted vertical line in the graph indicates the 3 km distance beyond which the WTP values of small and large choice sets are not significantly different for both intermediate and high ecological quality levels. Hence, the different choice set sizes do not result in significantly different WTP estimates for water quality changes in the study area and the hypothesis H_0^1 cannot be rejected.

5.2.2. Test 2: Swait and Louviere (SL) test

The second test of choice set effects is the SL test. This test is used to address the stability of preferences for the attributes (H_0^{1a} , H_0^{2a}) and possible scale differences (H_0^{1b} , H_0^{2b}) between different choice cards and subsets of choice tasks.¹⁵ The results of these tests are presented in Table 3, where rows 1–5 present the results for the comparison of small and large choice sets, whilst rows 6–8 provide the test results of the comparison between the first and second round with small choice sets. Data from the 759 respondents who completed all nine choice tasks were used.

The SL test results show that H_0^{1a} is rejected when the small and large datasets as a whole are compared (row 1). Because the scale and beta parameters are confounded, this result indicates that there is a significant difference in the scale or beta parameters between large and small choice sets, but it is impossible to attribute this to differences in either the beta and the scale parameters or the beta parameters alone. Further tests (rows 2 and 3) show that both the first round with small choice sets ($t=1,2,3$) as well as the second round ($t=8,9$) are significantly different from large choice sets. This could be caused by increasing the number of alternatives in the choice sets of the experiment, but also from learning of respondents about the choice task. We therefore further test the differences between small and large choice sets on a task-by-task basis. Similar results are found when testing H_0^{1a} and H_0^{1b} by comparing the first choice task with small choice sets to the first task with large choice sets ($t=1$, $t=4$; row 4).

¹³ Distance is included in the models after natural log transformation ($\ln(\text{km}+1)$), because this improved the model fit compared to linear, quadratic or reciprocal transformations. This transformation applies to all interaction terms of distance with other variables. The log transformation is one of the most commonly applied transformations in distance-decay studies for environmental valuation (Schaafsma, 2010).

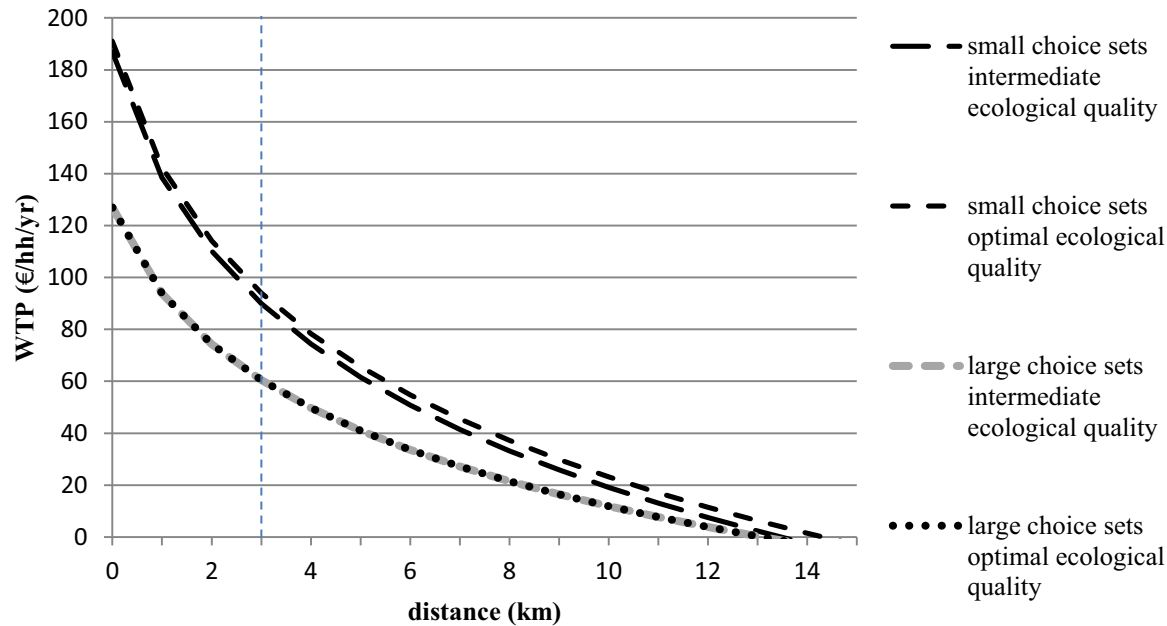
¹⁴ Two separate parameters for each of the quality levels in interaction with distance and the dummy variable were not significantly different and did not improve the model fit. Therefore, the variables were summed and the coefficients were thereby constrained to be identical.

¹⁵ In the estimation of the models to compare two choice tasks, the panel-data structure is not accounted for, as there is no panel effect, by definition, for one choice task, only when comparing two or more individual choice tasks by the same individual.

Table 2

WTP results for the small and large choice sets.

WTP	Model I: small choice sets	Model I: large choice sets
Intermediate ecological quality level	187 (167,209)	127 (105,152)
Optimal ecological quality level	191 (172,214)	127 (106,153)
Quality \times Distance	-70 (-79,-62)	-48 (-55,-41)
Intermediate ecological quality level at 3 km from the site	90 (70, 112)	60 (39, 85)



Notes: The 2.5% and 97.5% confidence intervals between brackets are calculated following [Krinsky and Robb \(1986\)](#), using 5000 draws. "Quality" refers to both ecological quality levels when included in the model in interaction with the "Distance" variable. The variable "Distance" is included in the model using a natural log transformation: $\ln(\text{km}+1)$. In the figure, the curves of intermediate and optimal quality of large choice sets overlap.

Table 3
Swait and Louviere test-results^a.

Row	H_0	(A)–(B)	LL(A)	LL(B)	LL(A&B) ($\mu_A \neq \mu_B$)	LR test (d.f. = 19) ^b	Reject H_0^a : $\beta_A = \beta_B$?	Rel.scale (μ_A/μ_B)	LL(A&B) ($\mu_A = \mu_B$)	LR test (df = 1) ^c	Reject H_0^b : $\mu_A = \mu_B$?
1	$H_0^{1a,b}$	Small sets–Large sets ($t=1,2,3,8,9-t=4,5,6,7$)	–3828	–3987	–8095	560	Yes	–	–	–	–
2		1st round small sets–Large sets ($t=1,2,3-t=4,5,6,7$)	–2823	–4578	–7424	48	Yes	–	–	–	–
3		Large sets–2nd round small sets ($t=4,5,6,7-t=8,9$)	–4578	–1878	–6478	43	Yes	–	–	–	–
4		Small sets 1st card–large sets 1st card ($t=1-t=4$)	–887	–1176	–2079	32	Yes	–	–	–	–
5	$H_0^{2a,b}$	Small sets 3rd card–large sets 1st card ($t=3-t=4$)	–929	–1176	–2112	16	No	0.87	–2113	2.1	No
6		1st round small sets–2nd round small sets ($t=1,2,3-t=8,9$)	2823	1878	4706	9	nNO	1.00	4706	0.0	No
7		Small sets 3rd card–small sets 4th card ($t=3-t=8$)	–929	–926	–1858	6	No	1.16	–1861	4.3	Yes
8		Small sets 3rd card–small sets 5th card ($t=3-t=9$)	–929	–948	–1882	10	No	1.00	–1882	0.3	No

^a Steps SL test: estimate two separate models (A) and (B) for (sets of) single choice tasks and register their log-likelihood; estimate a pooled model based on the two (sets of) choice tasks together (A&B) without restrictions on the scale parameter ($\mu_A \neq \mu_B$) by performing a grid search to optimise the log-likelihood under different relative scale adjustments for the second choice task (or set of tasks); perform a LR test for H_0^a QUOTE : $\beta_A = \beta_B$; estimated a pooled model keeping scale parameters constant for both (sets of) choice tasks ($\mu_1 = \mu_2$) and register the log-likelihood; perform a LR test for H_0^b QUOTE

^b χ^2_{19} ($p=0.05$)=30.1.

^c χ^2_1 ($p=0.05$)=3.84.

Only between the third task with choice small sets and the following first task with large choice sets ($t=3, t=4$), H_0^{1a} cannot be rejected (row 5). The subsequent test shows that adjusting the relative scale parameter for task $t=4$ does not significantly improve the log-likelihood of the model. Hence no scale parameter differences are found and H_0^{1b} can also not be rejected.

Next, the choices from small choice sets before and after exposure to large choice sets with more substitutes are compared. H_0^{2a} and H_0^{2b} are tested by comparing the first round with small choice set with the second round (row 6), as well as on a task-by-task basis by comparing the third choice task of the first round with small choice sets ($t=3$) to the two choices in the second round ($t=8, t=9$) (rows 7 and 8). In none of these tests significant changes in beta parameter estimates are found, so that H_0^{2a} cannot be rejected, which indicates that preferences for the attributes in the model may be stable. Subsequent tests for differences in the scale parameters show that H_0^{2b} cannot be rejected either when comparing the first and second round (row 6), i.e. adjusting the scale parameter does not improve the model fit compared with a model where an equality restriction is imposed. Only for the first choice in the second round with small choice sets directly following the large choice sets ($t=8$) the relative scale parameter is higher than for the third choice with small choice sets ($t=3$) and H_0^{2b} is rejected, but the difference is not highly significant. When comparing the third and the last choice with small choice sets ($t=3, t=9$), H_0^{2a} and H_0^{2b} cannot be rejected (row 8). Overall, this suggests that no net learning and fatigue effect is found and confirms the results of Model I. Hence, no evidence can be detected for preference instability in choices in small choice sets after consideration of large choice sets containing alternatives on a wider geographical scale.

5.2.3. Test 3: error-components Model II

The final tests of H_0^1 and H_0^2 are performed by estimating the pooled Model II. The estimates are presented in the last column in Table 1a. The error-components for large choice sets and the second round with small choice sets are included in the upper part of Table 1a under the random parameters, because they are included in the model as random parameters with zero mean and a normal distribution. This allows for controlling for correlation between the error-components and the random distance parameter. The random parameters in the model are assumed to be correlated and the off-diagonal elements of the parameter variance-covariance matrix are therefore allowed to be non-zero in model estimation. Cholesky decomposition (Train, 2003; Hensher et al., 2005a) is used to separate cross-correlation between random parameters (represented by the off-diagonal elements) from their individual standard deviations (represented by the diagonal elements). The lower triangular Cholesky matrix is presented in Table 1b.

The results in Table 1b show that, apart from the random distance and quality variable, only the error-component for large choice sets (λ_{in}^1) is significant. For Model II, the Cholesky decomposition reveals that the error-component for small choice sets is not significant (as can be seen from its diagonal value of the Cholesky matrix in Table 1b), after the significant correlation with the ($X \times D$) random parameter is controlled for (as can be seen from the off-diagonal value), and H_0^{2b} cannot be rejected. The result implies that the variance for large choice sets is different than for small choice sets and H_0^{1b} is

rejected. The variance increases with the larger number of alternatives in the choice sets, which can be attributed to higher choice complexity experienced by the respondents.

The model results suggest that when the difference in variance is controlled for, no differences in preferences for the attributes are observed. The insignificant interaction terms of the dummy for large choice sets and the quality attributes (β^L) indicate that the attributes do not have a different value in large choice sets compared to small choice sets. In other words, the number of alternatives does not affect the parameter estimates for the quality attributes and H_0^{1a} cannot be rejected. Furthermore, the error-component for the second round with small choice sets (λ_{in}^{Sb}) is not significant (see Table 1b), and hence no indication of higher variance signalling that respondents have learned or become fatigued, or a net effect of these factors. The interaction terms of the dummy for the second round with small choice sets (β^{Sb}) are also insignificant, indicating that preferences in the second round have not changed compared to the first round after the exposure to large choice sets. H_0^{2a} and H_0^{2b} cannot be rejected based on the results of this model.^{16,17}

In summary, the expansion of choice sets over a wider geographical area does not lead to significantly different WTP estimates for the quality levels at the lakes in the choice sets. The separate models for small and large choice sets show that there are no significant differences in WTP resulting from increasing the geographical scale of choice sets beyond 3 km from the sites. The results of the SL test indicate that there are some minor differences between the choices in small and large choice sets. The results of Model II indicate that the parameter estimates are similar, and only the variance differs between small and large choice sets: the expansion of choice sets is associated with higher variance. It may be that a lack of statistical power leads to making a Type II error. Nevertheless, the conclusion from the combined results of the three tests is that preferences may be stable, despite the higher variance in large choice sets. The results of this analysis hence suggest that respondents are able to evaluate seven alternatives (and an opt-out possibility) simultaneously. Increasing the geographical scale and choice set size does not result in significant differences in preference parameters between choice tasks in this case, at least, once respondents have evaluated a subset of these alternatives.

6. Conclusions and suggestions for further research

This study has aimed to provide more insight into the trade-off between choice set relevance and choice set complexity by analysing geographical framing effects. It contributes to further understanding of the effect of increasing the geographical scale, and thereby the size of choice sets, on choice task complexity and consequently WTP estimates for ecological quality changes. Such spatial scale effects are particularly important for environmental valuation, because many environmental goods and services are inherently spatially defined and differentiated. A typical characteristic of spatial choice studies is that the number of alternative sites providing similar recreational benefits or nature amenities is large. Valuation studies using SP methods face a trade-off in study design between including as many alternatives as possible to ensure relevance and adequately capture substitute effects, and choice task complexity related to overburdening respondents with too many choice options.

Via the experiment in this paper, we analyse changes in preferences within the same sample of respondents as a result of spatial differences in choice set framing. To this end, the size of the choice sets was changed during the sequence of choices in the CE. The CE started with a subset of the four lakes nearest to the home location of the respondent, followed by larger choice sets covering a wider geographical scale, before returning to the small subset. Choices between larger numbers of alternatives were expected to be more complex, potentially resulting in different choice behaviour. The framing effects were tested by estimating error-component models, performing Swait and Louviere tests, and comparing the resulting WTP estimates.

The results of these tests indicate that changing the geographical frame of choice sets has no significant effect on the WTP for the proposed environmental changes. Larger choice sets were associated with higher error variance suggesting higher choice task complexity. However, increasing the geographical scale and choice set size did not result in significant differences in preference parameters between choice tasks in this case study. Furthermore, evaluation of all study sites in large choice sets did not result in different WTP estimates or preference parameters compared to the choices in smaller choice sets. It may be that these findings reflect a lack of statistical power to detect the differences that we examined, but our results suggest that respondents can deal with multiple sites at a wider geographical scale as separate alternatives in the choice sets.

A possible explanation for the low impact of increasing the number of alternatives on WTP may be that the questions and information preceding the CE helped respondents sufficiently to reflect on their preferences for water quality changes at the eleven lakes in the full set of alternatives. All lakes in the full set of alternatives for which quality improvements were proposed were introduced before the CE began, and respondents were asked several questions addressing their preferences

¹⁶ Model II is re-estimated with only the error-component for large choice sets in view of the fact that the error-component for small choice sets was not significant. Including this error-component improves the model fit significantly ($LL = -7566$, compared to $LL = -7610$ for Model I without this error-component). However, this model does not lead to different WTP estimates and is therefore not presented here.

¹⁷ We estimated additional models which also included interactions with the remaining variables (price, distance, user and non-use interaction) and the CE subsets. It is impossible to include all parameters and their respective set-interactions simultaneously, because with the dummy-variables in the model there are insufficient degrees of freedom. Therefore, we performed a partial analysis, with three different models, using 500 Halton draws. The results of these models confirm the findings presented in the paper: only the error-component for large choice sets is significant.

and attitudes regarding all eleven lakes. These questions go far beyond merely presenting a glossary of attributes and their levels in the choice task instructions. Including such a glossary is common practice in most CE studies, but does not necessarily lead to different WTP estimates than studies that do not provide such a glossary (Bateman et al., 2008). The influence of survey elements preceding a CE, such as warm-up questions, deserves more attention in CE studies assessing the complexity of choice set designs and related learning and preference stability hypotheses (see e.g., Cai et al., 2011).

Another potential explanation is that asking respondents to choose first among a subset of the alternatives in order to familiarise them with the choice task may have reduced the complexity involved in choosing among alternatives in large choice sets. Further research is necessary to see if different conclusions will result from a different order of choice sets, starting with large choice sets for example. A limitation of the study design was that all respondents were presented small choice sets prior to large ones. If part of the sample had started with large choice sets, the hypothesis that choice task complexity can be mitigated by presenting subsets of alternatives first could have been tested, as it would have provided the opportunity to control for ordering effects.

The results suggest that respondents are capable of evaluating different combinations of seven alternatives out of a full set of eleven in a SP survey, at least if all alternatives are thoroughly introduced, respondents are asked for their experiences with and perceptions of these lakes before going through the CE and are first offered subsets of alternatives. Such a study design gives respondents the opportunity to learn about the choice alternatives before entering the full choice sets of the CE. This reduces the possibility that increasing the geographical scale of the choice sets leads to differences in choices and preference parameters. The results of this study suggest that careful survey design ensures that preference and choice stability are not compromised.

Acknowledgement

This study was carried out and financed under the EU DG Research F6 project AquaMoney (SSPI-022723) (www.aquamoney.org). We are grateful to Alison Gilbert and Alfred Wagtenonk for their input into the case study design.

References

- Adamowicz, W., Louviere, J., Williams, M., 1994. Combining revealed and stated preference methods for valuing environmental amenities. *Journal of Environmental Economics and Management* 26 (3), 271–292.
- Alpizar, F., Carlsson, F., Martinsson, P., 2001. Using Choice Experiments for Non-Market Valuation. Working Papers in Economics. Department of Economics, Göteborg University.
- Arentze, T., Borgers, A., Timmermans, H., DelMistro, R., 2003. Transport stated choice responses: effects of task complexity, presentation format and literacy. *Transportation Research Part E: Logistics and Transportation Review* 39 (3), 229–244.
- Arrow, K., Solow, R., Portney, P.R., Leamer, E.E., Radner, R., Schuman, H., 1993. Report of the NOAA Panel on contingent valuation. *Federal Register* 58 (10), 4601–4614.
- Banerjee, S., Murphy, J.H., 2005. The scope test revisited. *Applied Economics Letters* 12 (10), 613–617.
- Bateman, I.J., Burgess, D., Hutchinson, W.G., Matthews, D.I., 2008. Learning design contingent valuation (LDCV): NOAA guidelines, preference learning and coherent arbitrariness. *Journal of Environmental Economics and Management* 55 (2), 127–141.
- Bateman, I.J., Brouwer, R., Ferrini, S., Schaafsma, M., Barton, D.N., Dubgaard, A., Hasler, B., Hime, S., Liekens, I., Navrud, S., De Nocker, L., Ščeponavičiūtė, R., Semėnienė, D., 2011. Making benefit transfers work: deriving and testing principles for value transfers for similar and dissimilar sites using a case study of the non-market benefits of water quality improvements across Europe. *Environment & Resource Economics* 50 (3), 365–387.
- Ben-Akiva, M., Boccara, R., 1995. Discrete choice models with latent choice sets. *International Journal of Research in Marketing* 12 (1), 9–24.
- Boyle, K.J., Bergstrom, J.C., 2001. Doubt, doubts, and doubters: the genesis of a new research agenda?. In: Bateman, I.J., Willis, K.G. (Eds.), *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EU, and Developing Countries*, Oxford University Press, Oxford.
- Braga, J., Starmer, C., 2005. Preference anomalies, preference elicitation and the discovered preference hypothesis. *Environmental and Resource Economics* 32 (1), 55–89.
- Brouwer, R., 2008. The potential role of stated preference methods in the Water Framework Directive to assess disproportionate costs. *Journal of Environmental Planning and Management* 51 (5), 597–614.
- Brouwer, R., Dekker, T., Rolfe, J., Windle, J., 2010. Choice certainty and consistency in repeated choice experiments. *Environmental and Resource Economics* 46 (1), 93–109.
- Brown, T.C., Duffield, J.W., 1995. Testing part-whole valuation effects in contingent valuation of instream flow protection. *Water Resources Research* 31 (9), 2341–2351.
- Brownstone, D., Train, K.E., 1998. Forecasting new product penetration with flexible substitution patterns. *Journal of Econometrics* 89 (1–2), 109–129.
- Brownstone, D., Bunch, D.S., Train, K., 2000. Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles. *Transportation Research Part B—Methodological* 34 (5), 315–338.
- Cai, B., Cameron, T.A., Gerdes, G.R., 2011. Distal order effects in stated preference surveys. *Ecological Economics* 70 (6), 1101–1108.
- Campbell, D., Hutchinson, W.G., Scarpa, R., 2008. Incorporating discontinuous preferences into the analysis of discrete choice experiments. *Environmental and Resource Economics* 41 (3), 401–417.
- Carson, R.T., Groves, T., 2007. Incentive and informational properties of preference questions. *Environmental and Resource Economics* 37 (1), 181–210.
- Carlsson, F., Raun Mørkbak, M., Bøye Olsen, S., 2010. The First Time is the Hardest: A Test of Ordering Effects in Choice Experiments. Working Papers in Economics 470. Göteborg University, Department of Economics.
- Carson, R.T., Mitchell, R.C., 1995. Sequencing and nesting in contingent valuation surveys. *Journal of Environmental Economics and Management* 28 (2), 155–173.
- Causse, S., de Dios Ortuzar, J., Rizzi, L.I., Hensher, D.A., 2005. Assessing the influence of design dimensions on stated choice experiment estimates. *Transportation Research Part B: Methodological* 39 (7), 621–640.
- Colombo, S., Calatrava-Requena, J., Hanley, N., 2007. Testing choice experiment for benefit transfer with preference heterogeneity. *American Journal of Agricultural Economics* 89 (1), 135–151.
- CPB, Centraal Planbureau, 2010. Centraal Economisch Plan 2010. Den Haag.
- Cummings, R.G., Ganderton, P.T., McGuckin, T., 1994. Substitution effects in CVM values. *American Journal of Agricultural Economics* 76 (2), 205–214.

- DeShazo, J.R., Fermo, G., 2002. Designing choice sets for stated preference methods: the effects of complexity on choice consistency. *Journal of Environmental Economics and Management* 44 (1), 123–143.
- DeShazo, J.R., Cameron, T.A., Saenz, M., 2009. The effect of consumers' real-world choice sets on inferences from stated preference surveys. *Environmental and Resource Economics* 42 (3), 319–343.
- Ferrini, S., Fezzi, C., Day, B.H., Bateman, I.J., 2008. Valuing Spatially Dispersed Environmental Goods: A Joint Revealed and Spatial Preference Model to Consistently Separate Use and Non-Use Values. CSERGE Working Paper EDM 0803. CSERGE, University of East Anglia, UK.
- Fotheringham, A.S., O'Kelly, M.E., 1989. *Spatial Interaction Models: Formulations and Applications*. Kluwer Academic Print, Dordrecht.
- Haab, T.C., Hicks, R.L., 1997. Accounting for choice set endogeneity in random utility models of recreation demand. *Journal of Environmental Economics and Management* 34 (2), 127–147.
- Hailu, A., Adamowicz, W.L., Boxall, P.C., 2000. Complements, substitutes, budget constraints and valuation. *Environmental and Resource Management* 16 (1), 51–68.
- Hensher, D.A., Rose, J.M., Greene, W.H., 2005a. *Applied Choice Analysis: A Primer*. Cambridge University Press, Cambridge, UK.
- Hensher, D.A., Rose, J., Greene, W.H., 2005b. The implications on willingness to pay of respondents ignoring specific attributes. *Transportation* 32 (3), 203–222.
- Hoehn, J.P., Loomis, J.B., 1993. Substitution effects in the valuation of multiple environmental-programs. *Journal of Environmental Economics and Management* 25 (1), 56–75.
- Holmes, T.P., Boyle, K.J., 2005. Dynamic learning and context-dependence in sequential, attribute-based, stated-preference valuation questions. *Land Economics* 81 (1), 114–126.
- Jones, A., Wright, J., Bateman, I., Schaafsma, M., 2010. Estimating arrival numbers for informal recreation: a geographical approach and case study of British woodlands. *Sustainability* 2 (2), 684–701.
- Kahneman, D., 2002. Maps of Bounded Rationality: A Perspective on Intuitive Judgment and Choice, Nobel Prize Lecture, December 8, 2002.
- Kontoleon, A., Yabe, M., 2003. Assessing the impacts of alternative 'opt-out' formats in choice experiment studies: consumer preferences for genetically modified content and production information in food. *Journal of Agricultural Policy Research* 5 (5), 1–43.
- Kotchen, M.J., Reiling, S.D., 1999. Do reminders of substitutes and budget constraints influence contingent valuation estimates? Another comment. *Land Economics* 75 (3), 478–482.
- Krinsky, I., Robb, A.L., 1986. On approximating the statistical properties of elasticities. *Review of Economics and Statistics* 68 (4), 715–719.
- Loomis, J., Gonzalez-Caban, A., Gregory, R., 1994. Do reminders of substitutes and budget constraints influence contingent valuation estimates? *Land Economics* 70 (4), 499–506.
- Louviere, J.J., Hensher, D.A., Swait, J.F., 2000. *Stated choice methods: analysis and applications*. Cambridge, UK, Cambridge University Press.
- Louviere, J.J., Hensher, D.A., Swait, J.D., 2003. *Stated Choice Methods: Analysis and Applications*. Cambridge University Press, Cambridge.
- Manski, C.F., 1977. Structure of random utility models. *Theory and Decision* 8 (3), 229–254.
- Mas-Colell, A., Whinston, M.D., Green, J.R., 1995. *Microeconomic Theory*. Oxford University Press, New York, US.
- Mazzotta, M.J., Opaluch, J.J., 1995. Decision-making when choices are complex—a test of Heiners hypothesis. *Land Economics* 71 (4), 500–515.
- Neill, H.R., 1995. The context for substitutes in CVM studies: some empirical observations. *Journal of Environmental Economics and Management* 29 (3), 393–397.
- Ohler, T., Le, A., Louviere, J., Swait, J., 2000. Attribute range effects in binary response tasks. *Marketing Letters* 11 (3), 249–260.
- Parsons, G.R., Kealy, M.J., 1992. Randomly drawn opportunity sets in a random utility model of lake recreation. *Land Economics* 68 (1), 93–106.
- Parsons, G.R., Hauber, A.B., 1998. Spatial boundaries and choice set definition in a random utility model of recreation demand. *Land Economics* 74 (1), 32–48.
- Pate, J., Loomis, J., 1997. The effect of distance on willingness to pay values: a case study of wetlands and salmon in California. *Ecological Economics* 20 (3), 199–207.
- Payne, J.W., Bettman, J.R., Johnson, E.J., 1993. *The Adaptive Decision Maker*. Cambridge University Press, Cambridge, UK.
- Payne, J.W., Schkade, D.A., Desvousges, W.H., Aultman, C., 2000. Valuation of multiple environmental programs. *Journal of Risk and Uncertainty* 21 (1), 95–115.
- Pellegrini, P.A., Fotheringham, A.S., Lin, G., 1997. An empirical evaluation of parameter sensitivity to choice set definition in shopping destination choice models. *Papers in Regional Science* 76 (2), 257–284.
- Peters, T., Adamowicz, W.L., Boxall, P.C., 1995. Influence of choice set considerations in modeling the benefits from improved water quality. *Water Resources Research* 31 (7), 1781–1787.
- Puckett, S.M., Hensher, D.A., 2008. The role of attribute processing strategies in estimating the preferences of road freight stakeholders. *Transportation Research Part E—Logistics and Transportation Review* 44 (3), 379–395.
- Rolfe, J., Bennett, J., 2006. *Choice Modelling and the Transfer of Environmental Values*. Edward Elgar, Cheltenham, UK.
- Rose, J., Hensher, D.A., 2004. Modelling agent interdependency in group decision making. *Transportation Research Part E—Logistics and Transportation Review* 40 (1), 63–79.
- RuimteMonitor, 2008. Planbureau voor de Leefomgeving. (<http://www.ruimteMonitor.nl/>).
- Savage, S.J., Waldman, D.M., 2008. Learning and fatigue during choice experiments: a comparison of online and mail survey modes. *Journal of Applied Econometrics* 23 (3), 351–371.
- Scarpa, R., Ferrini, S., Willis, K., 2005. Performance of error component models for status-quo effects in choice experiments. In: Scarpa, R., Alberini, A. (Eds.), *Applications of Simulation Methods in Environmental and Resource Economics*, Springer, Dordrecht, the Netherlands.
- Schaafsma, M., 2010. *Spatial Effects in Stated Preference Studies for Environmental Valuation* (Ph.D. Thesis). VU University Amsterdam, the Netherlands.
- Scott, A., 2002. Identifying and analysing dominant preferences in discrete choice experiments: an application in health care. *Journal of Economic Psychology* 23, 383–398.
- Sutherland, R.J., Walsh, R.G., 1985. Effect of distance on the preservation value of water quality. *Land Economics* 61 (3), 281–291.
- Swait, J., Ben-Akiva, M., 1987. An analysis of the effects of captivity on travel time and cost elasticities. In: *Annals of the 1985 International Conference on Travel Behavior*. Noordwijk, Holland, pp. 113–128.
- Swait, J., Louviere, J.J., 1993. The role of the scale parameter in the estimation and comparison of multinational logit models. *Journal of Marketing Research* 30 (3), 305–314.
- Swait, J., Adamowicz, W.L., 1996. The effect of choice environment and task demands on consumer behavior: discriminating between contribution and confusion. *Staff Paper Series* 24091. University of Alberta.
- Swait, J., Adamowicz, W., 2001. The influence of task complexity on consumer choice: a latent class model of decision strategy switching. *Journal of Consumer Research* 28 (1), 135–148.
- Train, K., 2003. *Discrete Choice Methods with Simulation*. Cambridge University Press, New York.
- Tversky, A., 1972. Elimination by aspects: a theory of choice. *Psychological Review* 79 (4), 281–299.
- Whitehead, J.C., Blomquist, G.C., 1995. Do reminders of substitutes and budget constraints influence contingent valuation estimates: comment? *Land Economics* 71 (4), 541–543.